

Processing of native and foreign language subtitles in films: An eye tracking study

MARIE-JOSÉE BISSON, WALTER J. B. VAN HEUVEN, KATHY CONKLIN,
and RICHARD J. TUNNEY
University of Nottingham

Received: April 18, 2011 Accepted for publication: December 22, 2011

ADDRESS FOR CORRESPONDENCE

Correspondence concerning this article should be addressed to Marie-Josée Bisson, School of Psychology, University of Nottingham, Nottingham NG7 2RD, UK.
E-mail: lpxmb@nottingham.ac.uk

ABSTRACT

Foreign language (FL) films with subtitles are becoming increasingly popular, and many European countries use subtitling as a cheaper alternative to dubbing. However, the extent to which people process subtitles under different subtitling conditions remains unclear. In this study, participants watched part of a film under standard (FL soundtrack and native language subtitles), reversed (native language soundtrack and FL subtitles), or intralingual (FL soundtrack and FL subtitles) subtitling conditions while their eye movements were recorded. The results revealed that participants read the subtitles irrespective of the subtitling condition. However, participants exhibited more regular reading of the subtitles when the film soundtrack was in an unknown FL. To investigate the incidental acquisition of FL vocabulary, participants also completed an unexpected auditory vocabulary test. Because the results showed no vocabulary acquisition, the need for more sensitive measures of vocabulary acquisition are discussed. Finally, the reading of the subtitles is discussed in relation to the saliency of subtitles and automatic reading behavior.

An increasing number of films are imported from abroad and broadcast in the original foreign language (FL) soundtrack with subtitles added in the native language. This is called standard subtitling, and it is often preferred to dubbing as it is cheaper and keeps the original voice of the actors, thus avoiding the issue of lip synchronicity (Koolstra, Peeters, & Spinhof, 2002). Nowadays, it is often possible to add subtitles in different languages to films or television programs at the press of a button. When watching a film with subtitles, a viewer has to process not only three sources of information (the soundtrack, the subtitles, and the dynamic images in the film), but also the multilingual situation with both FL and native language. Furthermore, the information coming from these different sources may be redundant, which can render the reading of the subtitles less compelling.

Some studies in the 1980s investigated the allocation of attention to the different sources of information in this multimodal situation. They used eye tracking to measure the amount of time a viewer spent looking in the subtitle area as a function of the subtitle's presentation time with standard subtitles. In an initial study, D'Ydewalle, Muylle, and van Rensbergen (1985) found that participants fixated upon one or two words per subtitle, leading them to conclude that not much reading of the subtitles occurred. However, another study found that participants spent 30% of the subtitle's presentation time looking in the subtitle area (D'Ydewalle, van Rensbergen, & Pollet, 1987). Unfortunately, the descriptions of these two studies lack many methodological details (e.g., the number of participants, the language background of the participants, and the number of subtitles used in the study). In a study with children from Grades 2 to 6, D'Ydewalle and van Rensbergen (1989) found that the children also spent time looking in the subtitle area, and interestingly, this practice varied depending on the type of film: less time was spent in the subtitle area for an action film. The number of participants in that 1989 study was reported (12 in each experiment); however, the other methodological limitations mentioned above remain problematic. Overall, the first few studies on the processing of subtitles do indicate that viewers spend some of the presentation time looking at the subtitle area. However, because of the limitations of those studies, it is difficult to reach any more definitive conclusions.

One of the particular advantages of subtitling over dubbing is the potential for incidental acquisition of FL vocabulary.¹ In a study involving 246 primary schoolchildren watching a 15-min film with standard subtitles, Koolstra and Beentjes (1999) found that on a 28-item auditory vocabulary test, scores were higher in the group who had the FL soundtrack and the native language subtitles (20 correct answers) as opposed to the group with the FL soundtrack only (19 correct answers). These results must be treated with caution though because the control group, who were not exposed to the FL, were able to get 18 correct answers on the vocabulary test, and the analysis of variance (ANOVA) was within-subject despite it being a between-subject design. Therefore, it is difficult to conclude that there was vocabulary acquisition. Another study with primary schoolchildren found that scores on 10-item auditory and 10-item written vocabulary tests were higher in the condition with standard subtitles (5 correct answers on the auditory test and 5.8 correct answers on the written test) compared to a condition with both soundtrack and subtitles in the native language (4.1 correct answers on the auditory test, 4.8 correct answers on the written test; D'Ydewalle & van de Poel, 1999). However, this study used still images instead of a dynamic movie and found an advantage for standard subtitles only when the FL (Danish) used in the soundtrack was similar to the participant's native language (Dutch). Furthermore, the significance values for the post hoc comparisons are not provided in the article. While the results are promising, it would be important to replicate these findings with a longer vocabulary test and a movie with dynamic images before drawing any strong conclusions. The results of both these studies seem to indicate that subtitles are read to some extent and that the soundtrack is processed, thus allowing vocabulary acquisition to occur. However, it is also possible that the participants used the images in the film to form paired associations with the FL words in the soundtrack. It is therefore necessary to ascertain to what extent the

subtitles are read and to evaluate their usefulness in the process of FL vocabulary acquisition.

Another way of using subtitles is to add FL subtitles to a native language soundtrack. This is called reversed subtitling, and it was first investigated in the context of transcribed television and radio programs that did not involve visual images. Although in this case only two sources of information were available, it was concluded that reversed subtitling allowed participants to access meaning through the native language soundtrack and to map this onto the FL in the subtitles (Lambert & Holobow, 1984). In their study using still images, D'Ydewalle and van de Poel (1999) also investigated the acquisition of FL vocabulary using reversed subtitles. Setting aside the limitations mentioned earlier, they found that participants in the reversed subtitling condition (5.1 correct answers) outperformed participants in the control condition (4.1 correct answers) in the 10-item written vocabulary test, which seems to confirm that the FL subtitles were processed. However, a direct measure of the processing of the reversed subtitles using eye tracking would be more informative.

A more recent study using eye tracking investigated the reading of standard and reversed subtitles in children and adults using a 15-min animation. D'Ydewalle and De Bruycker (2007) report detailed eye movements of their participants on 114 and 138 subtitles for the standard and reversed subtitling conditions respectively. They found overall that reversed subtitles are skipped more often than standard subtitles (21% compared to 4%), are fixated less (0.59 fixations per word compared to 0.91) and that less time is spent in the subtitle area (26% of the subtitle's presentation time compared to 41%). Unfortunately, this study (and other prior studies, e.g., D'Ydewalle et al., 1985, 1987; D'Ydewalle & Van Rensbergen, 1989) did not include a control group with no subtitles, and therefore, the above values may reflect some time spent in the subtitle area as a result of other visual aspects of the movie taking place there. However, what seems to emerge from this study is that reading behavior does occur with both standard and reversed subtitles, but there seems to be a preference for native language subtitles.

Another method of subtitling presents both the soundtrack and subtitles in the same language, and this is referred to as intralingual subtitling. Originally intended to make films and TV programs accessible to the deaf and hearing-impaired community (Burnham et al., 2008; De Linde & Kay, 1999), intralingual subtitles are also used by language teachers and researchers because they have the potential to help the learner map the phonology to the written words. Written words in intralingual subtitling are not affected by intonation, accents, or background noise; thus, having access to both the soundtrack and subtitles allows for easier word segmentation by indicating which words are being spoken (Bird & Williams, 2002; Mitterer & McQueen, 2009). One study investigated the processing of intralingual native language subtitles while participants watched either a 12-min English or 20-min Dutch film (D'Ydewalle, Praet, Verfaillie, & van Rensbergen, 1991). The results showed that even though the native language subtitles were not necessary for the comprehension of the film, as a native language soundtrack was available, participants still read them approximately 20% of the time. The authors argued that participants read the subtitles because it was more efficient than listening to the soundtrack. However, as the same message was available in both soundtrack

and subtitles, it is highly possible that participants followed along with the written text while they heard the message aurally. Many vocabulary-learning studies seem to confirm that having both written and aural form of a word facilitates learning, which would imply that both are processed (Hu, 2008; Ricketts, Bishop, & Nation, 2009; Rosenthal & Ehri, 2008). It is crucial to note that these studies did not include dynamic images. Although studies of FL films with intralingual subtitles seem to show promising results in terms of improvement in speech performance (Borras & Lafayette, 1994), speech perception (Mitterer & McQueen, 2009), and word form-meaning associations (Sydorenko, 2010), indicating that both phonological and orthographic forms are processed, it is still necessary to investigate the reading of the subtitles when all three sources of information (FL soundtrack, FL subtitles, and dynamic images) are present.

Overall, the results of prior studies of incidental acquisition of FL vocabulary through watching films with subtitles provide some indirect evidence that the subtitles are processed. In addition, some studies have used eye tracking to directly measure the processing of the subtitles, but those studies are hampered by methodological issues or a lack of crucial information that makes them unreplicable. The aim of the current study was to use a film with standard, reversed, or intralingual subtitles to clarify the processing of the subtitles in each condition. Based on the previous studies mentioned above, it was predicted that participants would read the subtitles regardless of the subtitling condition, but that the reading behavior, as assessed by the number and duration of the fixations, would differ in each subtitling condition. More specifically, the duration and number of fixations should be higher in the standard condition as the subtitles are in the native language, followed by the intralingual condition where both subtitles and soundtrack are in a FL, and finally the reversed condition where the FL subtitles are superfluous and the FL is unknown to the participants. Although the main focus of the experiment was on the reading behavior, an auditory vocabulary test was used to assess the incidental acquisition of FL vocabulary in each subtitling condition.

Dutch was chosen as the FL in the film because most people in the United Kingdom have no knowledge of it. However, because Dutch comes from the same Germanic language family as English, the two languages have many similarities, which could aid the acquisition of Dutch vocabulary.

METHOD

Participants

Sixty-four participants took part in this experiment and received £6 for their participation. All participants completed a self-reporting language questionnaire to verify their language background and to ensure that they were native English speakers without any knowledge of Dutch. Because of the information reported on the language questionnaire, 10 participants had to be excluded for the following reasons: 2 participants had reading difficulties, 3 participants were not monolingual, 3 participants were not native English speakers, and 1 participant had spent 74 days in a Dutch-speaking country. The education level of the participants was varied, with 2 participants having completed at least secondary school, 26

currently at undergraduate level, and 18 at postgraduate level (8 participants did not provide this information). All participants were currently living in the United Kingdom.

In total, 36 participants (mean age = 24.6, 26 females) were included in the eye movement analysis, and 54 participants (mean age = 24.2, 39 females) were included in the vocabulary-test analysis. Even though 54 participants qualified for the experiment, 11 were in the “no movie” condition, and as such, no eye movement data was collected for them. Of the 43 participants in one of the movie conditions, eye tracking data was not recorded for 7 of them due to failed calibrations or software problems. When this was the case, participants still watched the film and took part in the vocabulary test.

Design

This experiment had a between-subject design with four conditions: (a) control condition: Dutch audio without subtitles (DA-NS); (b) intralingual subtitling: Dutch audio with Dutch subtitles (DA-DS); (c) standard subtitling: Dutch audio with English subtitles (DA-ES); and (e) reversed subtitling: English audio with Dutch subtitles (EA-DS). Another group of participants did not watch the movie and took part only in the vocabulary test.

Stimuli

The film excerpt lasted 25 min and consisted of four DVD chapters (chapters 2 to 5) of the movie *SpongeBob Square Pants* (Hillenburg, 2004). An animated film was chosen because the language of the soundtrack can be changed easily without affecting lip synchronicity. Four different versions of the movie were created and reformatted into avi files, which are compatible with the Experiment Builder software. The DA-NS version was created using the original movie chapters in Dutch. The DA-DS version was also created using the original movie chapters in Dutch, to which a transcription of the Dutch audio soundtrack was added as subtitles. It was necessary to create our own Dutch subtitles in order to have a true word-for-word transcription instead of the abbreviated subtitles that came with the movie DVD. We did, however, use the original subtitle timing information to add our own transcription to the movie using the subtitling program Submerge. The DA-ES version was created by using the original English subtitles and reformatting them using Submerge in order to match the DA-DS version with regards to the font, size, color, and display settings of the subtitles. The EA-DS version was created by adding our own Dutch subtitles to the English soundtrack, as with the previous two versions.

The vocabulary test consisted of two lists of 78 items. The items were one-word audio extracts selected from the movie soundtrack in Dutch. They varied in frequencies of occurrence in the movie (from 1 to 115 in the Dutch version of the movie and from 2 to 132 in the English version of the movie) and type of words (30 nouns, 16 pronouns, 10 adverbs, 6 adjectives, 6 verbs, and 10 other word types.) Each item was presented once with the correct English translations in the match trials (e.g., *koning*–king) and once with a foil in the mismatch trials (e.g.,

koning—coming). The foils were written English words that were phonologically similar to the Dutch ones and were taken from the movie's English subtitles wherever possible. In cases where no word in the English subtitles was phonologically similar, another English word, which was possible in the context of the movie, was chosen. This was done in order to keep the participants from relying solely on the phonological similarity or the context of the movie to complete the vocabulary test. The order of presentation of the lists was counterbalanced, and each item appeared once in each list in a random order. Each list contained half match and half mismatch trials such that if an item was presented with their correct translation in List 1, it was presented with its foil in List 2. The allocation of match and mismatch trials for each list was pseudorandomized.

Procedure

Participants were tested individually. They were informed that they would have to watch an animated film and were shown the head-mounted eye tracker they would wear during the film viewing. They were told that the film could be either in English or in a FL and that there might be subtitles. They were asked to watch the film as they would normally do at home and were told that there would be some questions to answer afterward about the film and about themselves. Participants were not explicitly asked to read the subtitles nor were they told to pay attention to the FL. Furthermore, they were not told that there would be a FL vocabulary test after the film.

The eye tracker was set up and calibrated using a 9-point calibration grid at the beginning of the session, and it was recalibrated after each chapter in the movie (four times total). The movie chapters were presented in the same sequential order for all participants so that they could follow the storyline. After watching the animation, participants completed an auditory vocabulary test and a language questionnaire.² In the vocabulary test, participants heard a one-word extract from the Dutch movie soundtrack three times while viewing either the correct English translations or a foil on the screen. They then had to decide whether the word they heard in Dutch had the same meaning as the English word they saw. The entire experiment lasted approximately 1 hr.

Participant's eye movements were recorded using an Eyelink 1 eye tracker (SR Research, Canada). This is a head-mounted eye tracker with a sampling rate of 250 Hz, equivalent to a temporal resolution of 4 ms. The eye tracker algorithm parses the eye movements into saccades (when movement of 0.5 degrees of visual angle or more is detected for two or more samples in a sequence), blinks (when the pupil data is missing for three or more samples in a sequence) and fixations (any period that is neither a saccade nor a blink), and records these events into an EDF file (Eyelink Data File) stored on the host computer. The movie chapters and vocabulary test were displayed on a PC using SR Research's Experiment Builder software and E-Prime, respectively.

RESULTS

The fixations were first split according to whether they occurred in the image area or the subtitle area. The subtitle area was taken as an area of 1024 (whole width of

screen) \times 218 pixels that started 50 pixels from the bottom of the display screen. This was deliberately larger than the actual area that displayed the subtitles to account for small vertical and horizontal inaccuracies in the recording of the eye movements. The image area included an area of 1024×450 pixels that started 50 pixels from the top of the display screen. The subtitle's timing information was used to determine whether a fixation occurred during the presentation of a subtitle.

The eye tracker output files were split into four chapters for each participant. Each movie chapter was then processed using the Eyetracker Output Utility program (van Heuven, 2010). The data from eight chapters (6% of the data) was excluded from the analysis because of excessive vertical drift, whereby all fixations are shifted up or down. Only chapters where this problem clearly occurred from the start were excluded. It was also necessary to discard some of the subtitles as they occurred where there was already some writing in English in the image or the subtitle area as part of the movie. This was the case for 10% of the Dutch subtitles and 8% of the English subtitles. Furthermore, the control condition, DA-NS, was used to exclude the subtitles for which other visual aspects of the movie co-occurred in the subtitle area. This involved a further 24% of Dutch subtitles and 20% of the English subtitles. In total, fixations occurring during 249 Dutch and 311 English subtitles were included in the analysis (see Table 1), which represents 66% and 71% of the original subtitles, respectively. Furthermore, Dutch subtitles were presented during 47% of the 25-min movie as opposed to 50% in the case of English subtitles. Subtitles comprised either one or two lines of text of different lengths and were presented on average for 3 s for the Dutch subtitles and 2.5 s for the English subtitles.

Subtitle area

The total fixation duration, the number of fixations, and the average fixation duration were calculated for each subtitle and averaged for each participant. The number of skipped subtitles (subtitles that participants did not fixate upon) was also calculated for each participant. All of these variables were then averaged for each condition (see Table 2) and submitted to a one-way ANOVA³ with subtitle condition as a between-subject factor (four levels: DA-NS, DA-DS, DA-ES, and EA-DS). Results revealed a significant main effect of subtitle condition for all four measurements: $F(3, 32) = 41.19, p < .001, \eta^2 = 0.79$ for the total fixation duration, $F(3, 32) = 49.98, p < .001, \eta^2 = 0.82$ for the number of fixations, $F(3, 32) = 5.49, p < .01, \eta^2 = 0.34$ for the average fixation duration and $F(3, 32) = 78.40, p < .001, \eta^2 = 0.88$ for the number of skipped subtitles. Tukey post hoc analyses revealed that participants spent longer fixating in the subtitle area when they heard the Dutch soundtrack with either Dutch or English subtitles than when they heard the English soundtrack ($p < .001$ in both cases). The pattern was exactly the same for both the number of fixations and the number of skipped subtitles. However, there were no significant differences between the condition with the Dutch soundtrack with Dutch subtitles and the Dutch soundtrack with English subtitles (means of total fixation duration, number of fixations, and skipped subtitles, all $ps > .84$). In addition, Tukey post hoc analyses revealed that the average fixation duration was significantly longer for the condition without subtitles than both conditions with subtitles and Dutch soundtrack (DA-DS, $p < .05$, and DA-ES, $p < .01$), but

Table 1. *Number of subtitles with exclusions, average subtitle duration, average number of words per subtitle, and number of subtitles with one and two lines of text for each chapter of the movie for Dutch and English subtitles*

Chapter	Dutch					English				
	1	2	3	4	Total	1	2	3	4	Total
All subtitles	108	76	102	89	375	128	89	121	98	436
Excluded										
English	13	7	4	12	36	13	6	4	13	36
Visual	45	15	19	11	90	40	16	20	13	89
<i>N</i> final	50	54	79	66	249	75	67	97	72	311
<i>n</i> 1 line of text	29	26	33	27	115	40	39	52	28	159
<i>n</i> 2 lines of text	21	28	46	39	134	35	28	45	44	152
<i>M</i> duration (ms)	2760	3093	3158	2759	2943	2410	2646	2568	2428	2513
<i>M</i> num. words	7.0	7.0	7.8	7.0	7.2	5.8	6.0	6.6	6.7	6.3

Table 2. *Mean (standard error) for the total fixation duration, number of fixations, average fixation duration in subtitle area and image area, and number of skipped subtitles for each subtitling condition*

Condition	<i>n</i>	Subtitle Area				Image Area		
		Tot. Fix Dur. (ms)	Num Fix.	Num. Skipped	Avg. Fix. Dur.	Tot. Fix Dur.	Num. Fix.	Avg. Fix. Dur.
		<i>M (SE)</i>	<i>M (SE)</i>	<i>M (SE)</i>	<i>M (SE)</i>	<i>M (SE)</i>	<i>M (SE)</i>	<i>M (SE)</i>
DA-NS (no subtitles)	8	81 (17)	0.28 (0.07)	186.88 (10.60)	283 (15)	2516 (54)	6.70 (0.32)	450 (36)
DA-DS (intralingual)	9	1285 (99)	5.65 (0.43)	20.78 (5.76)	240 (6)	1222 (88)	4.34 (0.33)	303 (18)
DA-ES (standard)	11	1284 (63)	5.90 (0.28)	10.82 (3.40)	227 (9)	832 (66)	3.16 (0.23)	274 (11)
EA-DS (reversed)	8	786 (135)	3.30 (0.54)	65.25 (15.24)	243 (10)	1738 (159)	5.56 (0.57)	350 (37)

Note: DA-NS, Dutch audio with no subtitles; DA-DS, Dutch audio with Dutch subtitles (intralingual subtitling); DA-ES, Dutch audio with English subtitles (standard subtitling); EA-DS, English audio with Dutch subtitles (reversed subtitling).

the difference was only approaching significance between the condition without subtitles and the condition with English soundtrack and Dutch subtitles ($p = .059$). The average fixation duration was not significantly different between the conditions with subtitles (all $ps > .68$). In order to confirm that the experimental groups were looking in the subtitle area because of the presence of the subtitles rather than because of other visual aspects of the movie taking place there, each subtitled condition was compared with the no-subtitles control group. Tukey post hoc analyses showed that participants spent significantly longer fixating in the subtitle area for each experimental group than for the control group ($p < .001$ for each comparison). This significant difference between the experimental and control groups was also found for the number of fixations and the number of skipped subtitles (all $ps < .001$).

As the duration of the subtitles' presentation differed between the conditions with Dutch and English subtitles, normalized total fixation durations in the subtitle area were calculated by dividing the total fixation duration in the subtitle area by the duration of the presentation for each subtitle. This was then averaged for each participant and condition (see Table 3) and submitted to a one-way ANOVA. Results showed that there was a main effect of condition, $F(3, 32) = 48.85$, $p < .001$, $\eta^2 = 0.82$. Tukey post hoc analyses revealed the same pattern as above, with every condition being significantly different from each other (all $ps < .01$) except for DA-DS and DA-ES ($p = .26$).

Similarly, because of the unequal number of words between the conditions with Dutch and English subtitles, normalized numbers of fixations were calculated by dividing the number of fixations by the number of words for each subtitle. This was then averaged for each participant and each condition (see Table 3) and submitted to a one-way ANOVA. Results indicated that there was a significant main effect of condition, $F(3, 32) = 56.42$, $p < .001$, $\eta^2 = 0.84$. Tukey post hoc analyses again revealed that differences between each condition were significant for the number of fixations even after controlling for the number of words in the subtitle (all $ps < .01$) except for the DA-DS condition and DA-ES that was again not significant ($p = .44$).

Furthermore, a normalized number of skipped subtitles was calculated by dividing the number of skipped subtitles by the total number of subtitles presented for each participant. This was then averaged for each condition and submitted to a one-way ANOVA. The results showed once more a main effect of condition, $F(3, 32) = 107.59$, $p < .001$, $\eta^2 = 0.91$. Tukey post hoc analyses continued to show the same pattern with the DA-DS and DA-ES conditions not being significantly different ($p = .59$). All other conditions were significantly different from each other (all $ps < .01$).

In an attempt to distinguish the reading behavior in the conditions with Dutch audio and either Dutch or English subtitles, the number of consecutive fixations in the subtitle area was calculated for each participant for each subtitle (these ranged from 2 to 20 consecutive fixations). The proportion of consecutive fixations was then calculated by adding up the number of consecutive fixations and dividing by the total number of fixations in the subtitle area. This process was repeated for each minimum number of consecutive fixations and averaged for each condition. For example, the proportion of fixations in the subtitle area for which there were 2

Table 3. *Mean (standard error) normalized total fixation duration, normalized number of fixations in subtitle area, and normalized number of skipped subtitles for each subtitling condition*

Condition	<i>n</i>	Normalized Total Duration			Normalized Fixations			Normalized Skipped Subtitles		
		<i>M</i> (<i>SE</i>)	Min	Max	<i>M</i> (<i>SE</i>)	Min	Max	<i>M</i> (<i>SE</i>)	Min	Max
DA-NS (no subtitles)	8	0.03 (0.01)	0.01	0.07	0.06 (0.01)	0.02	0.14	0.82 (0.03)	0.65	0.95
DA-DS (intralingual)	9	0.43 (0.03)	0.30	0.58	0.92 (0.06)	0.72	1.3	0.09 (0.03)	0	0.25
DA-ES (standard)	11	0.51 (0.03)	0.33	0.65	1.04 (0.04)	0.78	1.23	0.04 (0.01)	0	0.09
EA-DS (reversed)	8	0.27 (0.05)	0.11	0.46	0.59 (0.09)	0.26	1.03	0.26 (0.06)	0.02	0.47

Note: DA-NS, Dutch audio with no subtitles; DA-DS, Dutch audio with Dutch subtitles (intralingual subtitling); DA-ES, Dutch audio with English subtitles (standard subtitling); EA-DS, English audio with Dutch subtitles (reversed subtitling).

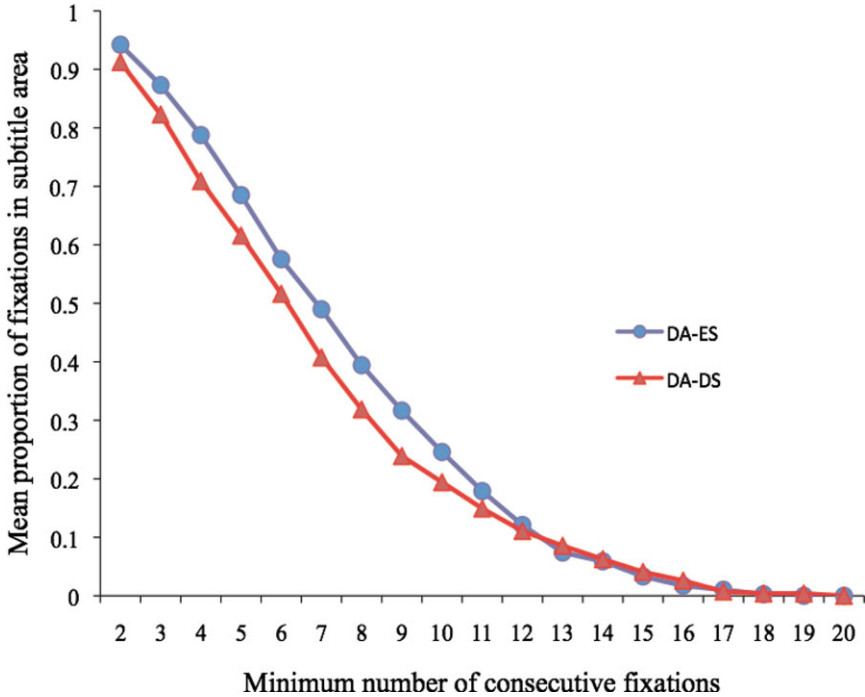


Figure 1. The mean proportion of fixations in the subtitle area that were consecutive for chapter 4 for the conditions with Dutch audio with English subtitles (DA-ES, standard subtitling) and Dutch audio with Dutch subtitles (DA-DS, intralingual subtitling). [A color version of this figure can be viewed online at <http://journals.cambridge.org/aps>]

or more consecutive fixations was .91 for the Dutch subtitle group and .94 for the English subtitle group, but the proportions dropped to .41 and .49 for 7 or more consecutive fixations for the Dutch and English subtitle groups respectively (see Figure 1). Chapter 4 was chosen for this analysis as it is the chapter with the biggest sample size and the most closely matched in terms of average number of words per subtitle between the Dutch and English subtitling conditions. An independent sample *t* test (two tailed) for each minimum number of consecutive fixations was calculated, and the results showed no significant differences between the two conditions for any minimum number of consecutive fixations (all *ps* > .17).

One of the reasons why participants may read the subtitles in the conditions with Dutch subtitles is the presence of Dutch–English orthographically similar or identical words. For example, the Dutch word “*kind*” is orthographically identical to the English word “kind” and the Dutch word “*promotie*” is orthographically similar to the English word “promotion.” Orthographically similar words can have different meanings if they are interlingual homographs (e.g., Dutch–English word “kind” meaning child in Dutch) or have the same meaning if they are cognates (e.g., “*promotie*”/“promotion”). To examine overlap in orthography regardless

of meaning, the number of Dutch words that were either identical or similar to English words in each subtitle was calculated using the normalized Levenshtein distance, which is a measure of orthographic similarity (Schepens, Dijkstra, & Grootjen, 2012).⁴ This was then correlated with the normalized number of fixations and the normalized fixation duration in the subtitles while controlling for the number of words in the subtitles (partial correlation). The result of the partial correlations showed no significant relationship between the number of identical or similar Dutch–English words and the normalized number of fixations for both the condition with Dutch soundtrack and Dutch subtitles ($r = -.02, p = .76$) and English soundtrack and Dutch subtitles ($r = -.05, p = .42$). Furthermore, there was no relationship between the number of identical or similar Dutch–English words and the normalized duration of fixations ($r = .10, p = .09$ for the condition with Dutch soundtrack and Dutch subtitles and $r = .05, p = .40$ for the condition with English soundtrack and Dutch subtitles).

Image area

The total fixation duration, the number of fixations, and the average fixation duration in the image area were calculated during each subtitle presentation and averaged for each participant and each condition (see Table 2) and submitted to a one-way ANOVA with subtitling condition as a between-subject factor (four levels). The results indicated a significant main effect of subtitling condition on the total fixation duration, $F(3, 32) = 58.49, p < .001, \eta^2 = 0.85$, the number of fixation, $F(33, 32) = 18.41, p < .001, \eta^2 = 0.63$, and the average fixation duration, $F(3, 32) = 9.11, p < .001, \eta^2 = 0.46$. Tukey post hoc analyses revealed that participants watching the film without subtitles spent more time fixating in the image area than the participants with the English soundtrack and Dutch subtitles condition ($p < .001$), and they in turn spent more time in the image area than participants with the Dutch soundtrack and Dutch subtitles ($p < .01$). However, the latter still spent more time in the image area than participants with the Dutch soundtrack and English subtitles ($p < .05$).

However, Tukey post hoc tests revealed that the number of fixations in the image area was significantly higher for the condition without subtitles compared to both conditions with Dutch audio and either Dutch or English subtitles (both $p < .001$), but this was not the case when comparing with the condition with English audio and Dutch subtitles ($p = .17$). There were also no significant differences between the conditions with Dutch audio and either Dutch or English subtitles ($p = .09$) or between the conditions with Dutch or English audio with Dutch subtitles ($p = .12$). However, the number of fixations in the image area was significantly higher in the condition with English soundtrack as opposed to English subtitles ($p < .001$).

In addition, Tukey post hoc tests showed that the average fixation duration in the image area was significantly longer for the condition without subtitles compared to the conditions with Dutch audio and Dutch or English subtitles ($p < .01$), but only marginally so for the condition with English audio and Dutch subtitles ($p = .059$). There were no significant differences between the conditions with subtitles (all $ps > .16$).

Vocabulary test

The percentage of correct responses on the vocabulary test for each participant was averaged for each condition. The reliability of the vocabulary test was high (Cronbach $\alpha = 0.75$). In the no-movie condition ($n = 11$), the mean percentage of correct responses on the vocabulary test was 60.7% ($SE = 2.2\%$). The mean score for the participants who watched the film with the Dutch soundtrack and no subtitles ($n = 9$) was 60.4 % ($SE = 2.0\%$). In the condition with Dutch soundtrack and Dutch subtitles ($n = 11$), the mean score was 61.5% ($SE = 1.4\%$). For the participants with Dutch soundtrack and English subtitles ($n = 13$), the mean percentage of correct answers on the vocabulary test was 62.4% ($SE = 2.4\%$). Finally, in the condition with English soundtrack and Dutch subtitles ($n = 10$), participants scored on average 60.1% ($SE = 3.0\%$). The mean scores on the vocabulary test were submitted to a one-way ANOVA, which revealed that there was no main effect of subtitling condition, $F(4, 49) < 1$. However, one-sample t tests showed that the mean scores for each condition were significantly higher than chance (all $ps < .001$).

DISCUSSION

The aim of the present study was to investigate viewers' reading of subtitles while they watched FL films with standard, reversed, and intralingual subtitles using eye tracking. One of the predictions was that when the soundtrack was in the FL, participants would spend more time reading the subtitles if they were in their native language as opposed to in the FL, but no significant differences were found for the duration of the fixations nor for the number of fixations in the subtitle area between those two conditions. Furthermore, no differences were found in the number of skipped subtitles between these two conditions.

The average fixation durations for the standard, intralingual, and reversed subtitling conditions (227, 240, and 243 ms, respectively) were in line with the average fixation duration (between 225 and 250 ms) for silent reading (Rayner, 1998, 2009). Fixations are normally slightly longer in the case of listening while reading (Rayner, 1998) which would have been expected at least with intralingual subtitles, as participants could have followed the FL words more closely in the subtitles while they heard them in the FL soundtrack. Another reason why longer average fixation durations might be expected in the case of the intralingual and reversed subtitling conditions is that participants were reading in an unknown FL. As this was the case, their average fixation durations were expected to be similar to the longer durations associated with low-frequency words, which normally reflect extra processing effort (Inhoff & Rayner, 1986; Rayner, 1998; Rayner & Raney, 1996). However, even though the average fixation durations were higher in both the intralingual and reversed subtitling conditions, these differences were not significant.

A possible explanation as to why we did not find differences in the average fixation durations is the lack of familiarity with the subtitling situation for the participants in the current experiment. In their study with adults and children from a subtitling country,⁵ D'Ydewalle and De Bruycker (2007) observed

shorter-than-average fixation durations in both the standard (179 ms) and reversed subtitling conditions (193 ms) for their adult participants. In the current study, fixation durations were similar to those obtained by the children in the D'Ydewalle and De Bruycker experiment (247 and 261 ms for the standard and reversed conditions, respectively), who did not have as much experience reading subtitles as the adults in the same experiment. It is unlikely that the participants in the current study watched subtitled films regularly,⁶ which could explain why their average fixation durations are similar to those of children in D'Ydewalle and de Bruycker (2007).

In the current study, even though participants in both the standard and intralingual subtitling conditions read most of the words in the subtitles, as indicated by the normalized number of fixations scores of about 1, they did not use the full subtitle presentation time to read the subtitles; instead they returned to view the image. This is even more evident in the intralingual condition where they could have benefited from a longer presentation time (because there were more words in the subtitles) to read the subtitles, but instead they used the extra available time to look at the image. This seems to support prior results indicating that the reading of subtitles still allows for the processing of the images (Perego, Del Missier, Porta, & Mosconi, 2010) instead of being just a reading exercise (Jensema, El Sharkawy, Danturthi, Burch, & Hsu, 2000).

The reading behavior in the standard and intralingual subtitling conditions was investigated further by looking at the consecutive fixations, and it was found that there were no significant differences between the two. This seems to indicate that participants were not merely attracted to specific words in the subtitles, but that once their gaze moved to the subtitle area they read in a normal, uninterrupted fashion instead of alternating between the image and the subtitle areas. Taken together, this seems to suggest that the reading behavior was similar across these two conditions. This is surprising since participants in the intralingual condition had no prior knowledge of the FL (Dutch) used in the subtitles. It is possible that participants were simply trying to use the available information in the subtitles to understand the movie since they could not rely on the FL soundtrack for understanding. On the other hand, participants in the reversed condition did not need the FL subtitles to help them understand the film because their native language was present in the soundtrack. However, they still spent a considerable amount of time in the subtitle area.

There may be different reasons why participants read subtitles. In view of the results of the current experiment, it seems unlikely that the appearance of subtitles on its own explains the reading behavior because the average normalized number of fixation was 0.92 for the intralingual condition (nearly 1 fixation per word) and 0.59 for the reversed condition (more than half of the words were fixated upon). The possibility that participants read the subtitles because of Dutch–English orthographically similar or identical words (cognates or interlingual homographs) was also ruled out as there was no relationship between the normalized number or duration of the fixations and the number of orthographically similar words. However, it is possible that the dynamic nature of the subtitles, i.e., the appearance and disappearance of the subtitles on the screen, coupled with the fact that the subtitles contained words was enough to generate the reading behavior. Moreover,

the automatic reading of words is well established (Laberge & Samuels, 1974; Samuels, 1994) and has an effect even in tasks such as the Stroop task in which reading is irrelevant (MacLeod, 1991; Stroop, 1935). However, it was surprising to find that this automatic reading behavior also seems to occur for FL words that were unknown to participants and that this reading occurred throughout the study. Both Dutch and English use an alphabetical script, which may have helped triggering the reading behavior. Furthermore, participants might have applied English spelling-to-sound rules to read the Dutch subtitles because of the similarities of the orthographic patterns between the two languages, which could have contributed to the reading behavior. Studies using languages with minimal orthographic (e.g., alphabet vs. nonalphabetic scripts) and phonological similarities could be used in the future to investigate this further.

Another way of explaining the reading of the subtitles is their saliency. Saliency maps have been used to understand the deployment of visual attention in static scene perception (Itti & Koch, 2000, 2001) and more recently in the area of dynamic scene perception (Mital, Smith, Hill, & Henderson, 2010). According to such research, the most salient feature attracts the viewer's gaze, which, in the case of dynamic scenes, is motion. Even when the goal of the viewer is oriented elsewhere, this salient feature will trigger an automatic saccade towards it. This supports the idea that participants may be attracted to the subtitle area first because of the dynamic nature of the subtitles. This might also explain why some subtitles were skipped: it is possible that subtitles are more likely to be skipped if their appearance on the screen co-occurs with a more salient feature in the image area. In addition, it has been shown that text in a visual scene is also salient and therefore likely to attract participant's gaze, even when it is not relevant to the task at hand (Cerf, Frady, & Koch, 2009), which could account for the reading of the subtitles even in the conditions where the subtitles are in FL. In future research, it would be interesting to use saliency maps with the different subtitling conditions in order to investigate the factors influencing the reading of the subtitles.

The reading of the subtitles might also be influenced by the redundancy of information in the image, the soundtrack and the subtitles. D'Ydewalle and De Bruycker (2007) observed that participants spent proportionally less time reading one-line subtitles compared to two-line subtitles in the standard subtitling condition. They argued that this is because some information in one-line subtitles is redundant as it is often already available in both the image and the soundtrack. As far as we know, the research on audiovisual redundancy has not yet looked at films with subtitles, and future studies could therefore investigate the influence of redundancy on the reading of subtitles.

Finally, there is also the possibility that participants read the subtitles because they believed this was expected of them, although they were not explicitly told to read subtitles. If this were the case, they presumably would have all read them regardless of the experimental condition. Our data shows, however, that participants read most of the subtitles in the intralingual and standard conditions, but only read some of the subtitles in the reversed condition. Thus, the assumption that they simply read the subtitles because they believed it was expected of them cannot account for the pattern of results.

The other aim of the current study was to investigate the incidental acquisition of FL vocabulary using an auditory vocabulary test. The mean percentages of correct responses on the vocabulary test were higher than chance, but as this was the case even when the movie was not watched, there is no evidence of vocabulary acquisition. The lack of differences across the conditions might be due to the limited exposure to the FL (25 min). However, because the vocabulary test measured knowledge at the recognition level only, it is possible that the participants did acquire some vocabulary knowledge, but that it did not reach the recognition level. Future studies should use a more sensitive measure of vocabulary acquisition. Alternatively, as the incidental acquisition of vocabulary is a slow process with small vocabulary gains (Nation, 2001; Paribakht & Wesche, 1997; Schmitt, 2010), the impact of long-term exposure to FL films with subtitles on language acquisition would perhaps be easier to assess, and therefore a longitudinal study might be more appropriate.

To summarize, the results indicate no significant differences between the standard and the intralingual subtitling conditions in terms of the fixation duration, the number of fixations, the number of skipped subtitles, or the number of consecutive fixations in the subtitle area, while participants spent longer looking at the image in the intralingual condition. Participants spent less time reading the subtitles in the reversed subtitling condition compared to the standard and intralingual subtitling conditions, but they did spend some time reading the subtitles even though they were not required to. It is probable that participants are attracted to the subtitle area primarily because of the saliency of the subtitles: their dynamic nature as well as the text they contain makes them very salient features. In addition, the high normalized number of fixations, even in the conditions with FL subtitles, indicate that reading behavior does occur. The possibility of using English spelling-to-sound rules as well as the automatic reading of text are most likely responsible for triggering the reading behavior in the FL subtitling conditions. Because subtitles were processed to some extent in each condition, it would suggest that vocabulary acquisition using FL films with subtitles is possible. However, it will be necessary to develop more sensitive measures of vocabulary acquisition or resort to longitudinal studies to answer this question. As language learners can also use the soundtrack to increase their knowledge of a FL, future studies will need to investigate to what extent the soundtrack is also processed when subtitles are available. Our current study showed that both the subtitles and the images in the FL films are processed, but further investigations that more explicitly investigate all three input modalities are needed in order to explore their effects on language acquisition.

NOTES

1. We use the expression “vocabulary acquisition” in this article to refer to an increase in knowledge about a lexical item. For an overview of the different aspects of word knowledge required to fully master a lexical item see Nation (2001) and Schmitt (2010).
2. All participants completed the language questionnaire at the end of the experiment. As the questionnaire enquired about language background and specifically asked whether participants had any prior knowledge of the FL used in the film (Dutch) or spent

- time in a country where Dutch is spoken, it was completed at the end of the experiment to avoid drawing attention to the FL vocabulary acquisition aspect of the study.
3. Each dependant variable was first submitted to a mixed-design ANOVA with condition as a between-subject factor (four levels: Dutch soundtrack with Dutch subtitles, Dutch soundtrack with English subtitles, Dutch soundtrack with no subtitles, English soundtrack with Dutch subtitles), and chapter as a within-subject factor (four levels: chapter 1 to chapter 4) in order to rule out an effect of chapter or an interaction between chapter and condition. Because an interaction between chapter and condition was found only for the number of skipped subtitles and this effect was driven by the control condition (Dutch audio with no subtitles), the dependent variables were collapsed across chapters for all other analyses.
 4. The normalized Levenshtein distance scores range from 0 (*orthographically identical*) to 1 (*completely different*). The criteria used to classify the words in this study was <0.3 .
 5. Subtitling countries in Europe include Belgium, Denmark, Finland, Greece, Luxembourg, The Netherlands, Portugal, and Sweden, as opposed to dubbing countries such as Austria, France, Germany, Italy, and Spain (Koolstra, Peeters, & Spinhof, 2002).
 6. Although we did not ask our participants about their audiovisual watching habits, most of them are living in the United Kingdom, which does not often use subtitles, so it is likely that they do not have as much experience with subtitles as someone living in a subtitling country.

REFERENCES

- Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, 23, 509–533.
- Borras, I., & Lafayette, R. C. (1994). Effects of multimedia courseware subtitling on the speaking performance of college-students of French. *Modern Language Journal*, 78, 61–75.
- Burnham, D., Leigh, G., Noble, W., Jones, C., Tyler, M., Grebennikov, L., et al. (2008). Parameters in television captioning for deaf and hard-of-hearing adults: Effects of caption rate versus text reduction on comprehension. *Journal of Deaf Studies and Deaf Education*, 13, 391–404.
- Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, 9, 1–15.
- D'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist*, 12, 196–205.
- D'Ydewalle, G., Muylle, P., & van Rensbergen, J. (1985). Attention shifts in partially redundant information situations. In R. Groner, G. W. McConkie, & C. Menz (Eds.), *Eye movements and human information processing* (pp. 375–384). North-Holland: Elsevier.
- D'Ydewalle, G., Praet, C., Verfaillie, K., & van Rensbergen, J. (1991). Watching subtitled television: Automatic reading behavior. *Communication Research*, 18, 650–665.
- D'Ydewalle, G., & van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research*, 28, 227–244.
- D'Ydewalle, G., & van Rensbergen, J. (1989). Developmental studies of text-picture interactions in the perception of animated cartoons with text. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 233–248). North-Holland: Elsevier.
- D'Ydewalle, G., van Rensbergen, J., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In J. K. O'Regan &

- A. Levy-Schoen (Eds.), *Eye movements from physiology to cognition* (pp. 313–321). Amsterdam: Elsevier.
- De Linde, Z., & Kay, N. (1999). Processing subtitles and film images: Hearing vs deaf viewers. *Translator*, 5, 45–60.
- Hillenburg, S. (Director). (2004). *Spongebob Squarepants The Movie* [Animated motion picture]. Nickleodeon.
- Hu, C.-F. (2008). Use orthography in L2 auditory word learning: Who benefits? *Reading and Writing*, 21, 823–841.
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, 40, 431–439.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2, 194–203.
- Jensema, C. J., El Sharkawy, S., Danturthi, R. S., Burch, R., & Hsu, D. (2000). Eye movement patterns of captioned television viewers. *American Annals of the Deaf*, 145, 275–285.
- Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development*, 47, 51–60.
- Koolstra, C. M., Peeters, A. L., & Spinhof, H. (2002). The pros and cons of dubbing and subtitling. *European Journal of Communication*, 17, 325–354.
- Laberge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.
- Lambert, W. E., & Holobow, N. E. (1984). Combinations of printed script and spoken dialogue that show promise for students of a foreign-language. *Canadian Journal of Behavioural Science–Revue Canadienne Des Sciences Du Comportement*, 16, 1–11.
- MacLeod, C. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163–203.
- Mital, P. K., Smith, T. J., Hill, R., & Henderson, J. M. (2010). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, 3, 1–15.
- Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *Plos One*, 4, A146–A150.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Paribakht, S. T., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. N. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174–200). Cambridge: Cambridge University Press.
- Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, 13, 243–272.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457–1506.
- Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, 3, 245–248.
- Ricketts, J., Bishop, D. V., & Nation, K. (2009). Orthographic facilitation in oral vocabulary acquisition. *Quarterly Journal of Experimental Psychology*, 62, 1948–1966.
- Rosenthal, J., & Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology*, 100, 175–191.
- Samuels, S. J. (1994). Toward a theory of automatic information processing in reading, revisited. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 816–837). Newark, DE: International Reading Association.
- Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distribution of cognates in Europe as based on the Levenshtein distance. *Bilingualism: Language and Cognition*, 15, 157–166.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Houndmills: Palgrave Macmillan.

- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14, 50–73.
- van Heuven, W. J. B. (2010). Eyetracker Output Utility (Version 1.20 beta 14) [Computer program]. Retrieved from <http://www.psychology.nottingham.ac.uk/staff/wvh/eou/index.html>